

ACCEPTABILITY OF ALTERNATIVE TREATMENTS FOR DEVIANT CHILD BEHAVIOR

ALAN E. KAZDIN

THE PENNSYLVANIA STATE UNIVERSITY

The acceptability of alternative treatments for deviant child behavior was evaluated in two experiments. In each experiment, clinical cases were described to undergraduate students along with four different treatments in a Replicated Latin Square Design. The treatments included reinforcement of incompatible behavior, time out from reinforcement, drug therapy, and electric shock and the treatments were described as they were applied to children with problem behaviors. Experiment 1 developed an assessment device to evaluate treatment acceptability and examined whether treatments were rated as differentially acceptable. Experiment 2 replicated the first experiment and examined whether the severity of the presenting clinical problem influenced ratings of acceptability. The results indicated that treatments were sharply distinguished in overall acceptability. Reinforcement of incompatible behavior was more acceptable than other treatments which followed, in order, time out from reinforcement, drug therapy, and electric shock. Case severity influenced acceptability of alternative treatments with all treatments being rated as more acceptable with more severe cases. However, the strength of case severity was relatively small in relation to the different treatment conditions themselves which accounted for large portions of variance.

DESCRIPTORS: social validation, treatment acceptability, punishment

Evaluation of behavior modification and more traditional psychotherapy techniques has relied almost exclusively on outcome measures of client change. The focus on outcome measures is of obvious significance because of the primary goal of developing effective treatment techniques. Recently, investigators have discussed the need for broader criteria to evaluate treatment, in addition to measures of efficacy (Garfield, 1978; Kazdin & Wilson, 1978; Strupp & Hadley, 1977; Wolf, 1978). Some of the proposed criteria to broaden treatment evaluation include the efficiency of treatment, side effects of treatment, discomfort and stress during treatment,

professional training required to administer treatment, cost effectiveness, and others.

One criterion recently proposed as an important dimension is the *acceptability* of treatment (Wolf, 1978). Acceptability refers to the judgments about the treatment procedures by non-professionals, lay persons, clients, and other potential consumers of treatment. Judgments of acceptability are likely to embrace evaluation of whether treatment is appropriate for the problem, whether treatment is fair, reasonable, and intrusive, and whether treatment meets with conventional notions about what treatment should be. In general, acceptability refers to the overall evaluation of the procedures.

Several reasons exist for evaluating the acceptability of treatment. First, for selected clinical problems, several techniques are available that effectively alter behavior. For example, in child treatment, variations of reinforcement, time out from reinforcement, response cost, overcorrection, and many other techniques alone

This research was supported by a grant (MH-31047) by the National Institute of Mental Health and the Central Fund for Research from The Pennsylvania State University. The author is grateful to Paul Games for statistical consultation with the data analyses. Requests for reprints should be sent to the author, Department of Psychology, The Pennsylvania State University, University Park, Pennsylvania 16802.

or in combination have been shown to suppress a variety of inappropriate behaviors (e.g., O'Leary & O'Leary, 1976; Sulzer-Azaroff & Mayer, 1977). Similarly, in adult outpatient treatment, variations of desensitization, flooding, modeling, and reinforced practice all have been shown to reduce markedly or to eliminate avoidance behaviors (e.g., Leitenberg, 1976; Marks, 1978). Presumably, treatments that are effective may vary considerably in their acceptability to consumers. Indeed, the most effective technique may not be the most acceptable to clients (see Crowe, Marks, Agras, & Leitenberg, 1972). Treatments viewed by the public as more acceptable than others are more likely to be sought by potential consumers, initiated, and adhered to once they are initiated based upon criteria other than efficacy alone.

A second reason to evaluate treatment acceptability pertains to ethical and legal issues that treatment procedures often raise. Litigation in the treatment of psychiatric patients, the mentally retarded, and prison populations has ruled that certain procedures, independently of their effects on behavior, are unacceptable because they infringe upon client rights (see Budd & Baer, 1976; Kazdin, 1978; Martin, 1975). To help protect client rights, institutional review committees and guidelines for treatment procedures have recommended including lay persons to evaluate whether the treatment procedures are acceptable in light of the client's problem (e.g., May, Risley, Twardosz, Friedman, Bijou, & Wexler, 1976). Thus, the acceptability of treatment has increased in importance as a treatment criterion in its own right.

Occasionally, published reports have included client evaluations of treatment procedures by mentioning anecdotal information about how institutional staff, parents, and clients prefer one treatment technique over another (Drabman, Spitalnik, & Spitalnik, 1974; Foxx & Azrin, 1972, 1973). In some cases, questionnaire data have been obtained about staff or client satisfaction with treatment (Foxx & Shapiro, 1978; Liberman, Levine, Wheeler, Sanders, & Wal-

lace, 1976; Porterfield, Herbert-Jackson, & Risley, 1976; Rosenbaum, O'Leary, & Jacob, 1975; Webster & Azrin, 1973). Typically, client preference and satisfaction are assessed by a few questions derived from considerations of "face validity" alone. Also, the differential efficacy of alternative procedures is often confounded with ratings of preference or satisfaction because treatments often vary markedly in what they achieve and because retrospective questions of acceptability of the procedures may reflect these different outcomes (e.g., Porterfield et al., 1976; Webster & Azrin, 1973). Additional research is needed to develop an assessment device to distinguish treatments independently of efficacy and that can be justified on more than considerations of face validity alone.

Presumably, an area in which great concern might be voiced about the acceptability of treatment pertains to the use of techniques to suppress behavior. Aversive techniques such as shock and isolation raise issues about client rights and whether the means of treatment are appropriate or acceptable. Punishment is an area of particular interest from a clinical standpoint because several different techniques have been shown to suppress behavior. A major issue for such techniques with demonstrated efficacy is their acceptability to parents, teachers, clients, and other possible consumers.

The present investigations examined the acceptability of different treatment procedures used to suppress deviant child behavior. The techniques include reinforcement of incompatible behavior, time out from reinforcement, drug therapy, and electric shock. The treatments were evaluated in relation to the descriptions of applications with behavior problem children. Treatment of children was selected for several reasons. First, problematic behavior of children has represented the largest single treatment focus of many behavioral techniques (Kazdin, 1975). Second, children were selected because they may raise major issues about treatment acceptability. Consumers of treatment are likely to be especially protective of children who are

often not considered to be competent to weigh the manifold considerations (e.g., risks and benefits) that enter into treatment decisions. Alternative treatments, as applied to childhood behavior problems, were evaluated to develop a methodology to assess treatment acceptability, to evaluate whether a select number of treatments with demonstrated efficacy are differentially acceptable, and to evaluate variables that may influence evaluations of acceptability.

EXPERIMENT 1

Overview

The first experiment examined whether alternative treatment procedures used for the treatment of deviant child behavior were differentially acceptable. Undergraduate students rated different treatment descriptions as they were applied to one of two behavioral problem children. After each description of a treatment, the student completed questionnaires to evaluate characteristics of that treatment. Four different treatments were presented in a 4×4 Replicated Latin Square Design in which each student rated each of the treatments in one of four sequences.

METHOD

Participants

Participants were 68 female and 20 male undergraduates recruited from introductory psychology courses in which course credit was given for participation. As students arrived individually for the experiment, they were assigned randomly to one of eight conditions (2 child cases \times 4 sequences) based upon a random ordering of conditions.

Assessment

Acceptability measure. Evaluating the acceptability of treatment required developing a measurement device. Treatment acceptability was considered to reflect a person's overall evaluation of the procedure including such dimensions as whether the treatment would be recom-

mended or endorsed for broad application; whether it was unfair or cruel; whether it would be appropriate, if applied to someone who was not capable of giving consent; and whether treatment was consistent with commonly held notions of what treatment should be. Forty-five items were generated that appeared on "face validity" to be related to client evaluation of treatment. Sixteen of these were selected because of their apparent relevance to treatment with children and the use of punishment. (Other items focused on dimensions of treatment more relevant to outpatient psychotherapy.)

The measure of acceptability, referred to as the Treatment Evaluation Inventory, required students to rate the 16 items in a Likert-type format (1- to 7-point scale). The item contents are illustrated by questions asking students to rate how acceptable treatment was, how willing they would be to carry out the procedure, how suitable the procedure would be for children with problems other than those described in the study, how cruel or unfair treatment was, and how much the student liked the procedure. The items were evaluated in pilot work by administering the questionnaire to 60 students (30 females, 30 males) in introductory psychology who heard one of four treatments as applied to a clinical case (see Treatment Conditions below). In addition to the items from the Treatment Evaluation Inventory, pilot students rated 15 bipolar adjectives from the Semantic Differential (Osgood, Suci, & Tannenbaum, 1957). The items covered the Evaluative, Potency, and Activity dimensions.

The individual item responses to the Treatment Evaluation Inventory and Semantic Differential were subjected to factor analysis. The purpose was to look for items from the Treatment Evaluation Inventory that would load consistently on a single dimension. The Semantic Differential items were included for two reasons. First, the items would increase the number of variables in the overall factor analysis and introduce different factors that would allow for a more careful delineation of a homogeneous fac-

tor for the Treatment Evaluation Inventory. Second, the Evaluative dimension of the Semantic Differential could provide partial validation of the Treatment Evaluation Inventory. Acceptability of treatment is likely to be related to the Evaluative dimension which has accounted for the major portion of variance in Semantic Differential research (Osgood et al., 1957). The Evaluative dimension, as represented by judgments of items related to "good-bad," reflects an overall positive or negative reaction. Items from the Treatment Evaluation Inventory designed to measure acceptability would be expected to correlate with Evaluative scores from the Semantic Differential. In general, the purpose of pilot work was to examine the items from the Treatment Evaluation Inventory.

Responses of the 60 pilot students to the Treatment Evaluation Inventory and Semantic Differential were subjected to a principal components factor analysis (Harman, 1976). The results indicated that 15 of the 16 items of the Treatment Evaluation Inventory produced high loadings on a single principal component before rotation (range from .67 to .94) and on the first factor after varimax rotation (range from .61 to .95). The first principal component on which these items loaded accounted for 51.4% of the variance. In the rotated factor analysis, items from the inventory loaded highly on a single factor and with the exception of one item did not load highly (loadings less than .40) on other factors. (One item loaded .61 on the first overall factor and .52 on a second factor.) Interitem correlations for items of the first factor of the Treatment Evaluation Inventory ranged from .35 to .96 (median $r = .67$).

As expected, loadings from the Evaluative dimension of the Semantic Differential also were high for this first factor suggesting that the Treatment Evaluation Inventory in fact assessed an overall evaluative reaction of the students.¹ The loadings for Potency and Activity dimensions of the Semantic Differential were low (less than .40) for the single factor that characterized items of the Treatment Evaluation Inventory.

One item on the Treatment Evaluation Inventory had a small loading (.24) on the overall single factor and was deleted from the final questionnaire. (This item asked students to rate the extent to which treatment was manipulative and coercive.) The 15-item Treatment Evaluation Inventory was developed from the above pilot work and was used in subsequent experiments, described below.²

Dependent measures. The dependent measures consisted of the Treatment Evaluation Inventory and the Semantic Differential.³ From the Semantic Differential, bipolar adjectives were selected for Evaluative, Potency, and Activity dimensions. Five items for each dimension were included. Characteristic items included good-bad, kind-cruel (Evaluative), strong-weak, heavy-light (Potency), and active-passive, fast-slow (Activity).

The Semantic Differential was included as a dependent measure for two reasons. First, the

¹The Treatment Evaluation Inventory was not combined into a single overall measure with the Semantic Differential Evaluative items for two reasons. First, the purpose was to develop a separate measure of acceptability for subjects to evaluate treatment that could be used independently of other scales such as the Semantic Differential. Second, the Treatment Evaluation Inventory and Semantic Differential consist of different rating formats.

²Since completion of the experiments of the present report, additional data have been obtained for the Treatment Evaluation Inventory. The Inventory and items from the Semantic Differential (same as those used in the present investigations) were administered to 144 additional subjects, college students enrolled in Introductory Psychology courses. They rated child treatment applications including some of the treatments in the present report and others, following the procedures described for administering the conditions in the present report. The results were factor analyzed as the original pilot data using a principal components factor analysis. Items from the Treatment Evaluation Inventory yielded high factor loadings on the first unrotated factor (range .56 to .91) and items from the Evaluative dimension of the Semantic Differential loaded highly (range .69 to .89) on this factor as well. These results support the original analyses using a separate and larger sample of subjects.

³Copies of the extended case descriptions, treatment descriptions, and assessment devices can be obtained from the author.

Semantic Differential includes the Evaluative dimension, as mentioned earlier, which may be related to acceptability of treatment. Because the Semantic Differential does not ask specific questions about treatment and is in a different rating format from the Treatment Evaluation Inventory, it provides a methodologically distinct assessment device to examine evaluative reactions (cf. Campbell & Fiske, 1959).

Another reason for using the Semantic Differential is that the measure permits examination of other dimensions that may be related to treatment. Specifically, Potency and Activity are assessed that reflect judgments about the strength of the treatment procedures. Judgments about how potent or active specific techniques are may or may not be related to overall acceptability of treatment. Intrusive or relatively extreme techniques (e.g., shock) may be rated as "strong" types of interventions and, perhaps, unacceptable for that reason alone. In any case, Potency and Activity provide additional details about the acceptability of treatment by elaborating possible considerations that go into such ratings.

Procedures

Participants who arrived for the experiment received instructions about the purpose of the experiment, namely, to evaluate different clinical treatments, and how to complete the questionnaires. Students received a packet of questionnaires that included four sets of the dependent measures, mentioned earlier. One set was used to evaluate each treatment. Each set of questionnaires in the packet was separated by a blank sheet with instructions to stop before proceeding further in the packet. Students were instructed by the experimenter not to look ahead nor to look back to previous responses when completing the questionnaires. Compliance with these instructions was monitored by the experimenter in an adjacent room during the experiment.

After the initial instructions, the experimenter left the room and played cassette tapes that presented the different treatments. The tapes played

in the room where the participant was seated. The procedures were monitored by the experimenter in an adjoining room separated by a one-way mirror. The tape recorder was controlled by the experimenter in such a way that the student could complete the questionnaires after hearing each treatment. After completing the questionnaires for a particular treatment, the student was instructed to look up to cue the experimenter to play the next description. An initial tape described the case of a problem child who was brought to treatment (see Treatment Conditions). After the case description each treatment was described as applied to that particular case.

Case Descriptions

The purpose of the initial study was to evaluate four different treatments used for treatment of behavioral problems. The treatments included reinforcement, time out from reinforcement, drug therapy, and electric shock. The treatments were evaluated by the participants after hearing one of two cases of children whose behavior warranted treatment. Two different case descriptions were used to assess the possibility that treatment evaluations of the students were based upon, or restricted to, unique characteristics of the stimulus material (cf. Maher, 1978).

The first description depicted a 5-year-old girl named Ann of normal intelligence. Ann's problems generally consisted of oppositional behavior at home. She was described as never following parental instructions related to such activities as getting dressed, doing chores, or coming when asked. Also, she was said to defy her parents purposely, to get mad at her younger brother for no apparent reason, and to whine constantly when she did not get her way. The parents were said to be concerned about Ann, particularly since she was about to enter kindergarten where her problems might create difficulties in the classroom.

The second description depicted a 10-year-old boy, named Ralph, who was educably mentally retarded. Ralph was described as having problems in his special education classroom. Specifi-

cally, Ralph was described as not doing what he was supposed to in class, not concentrating, and generally engaging in disruptive behavior. The teacher sought help because of the consistent disruptive behavior that Ralph had shown.

The two cases varied the sorts of problems that were as described. Both oppositional behavior at home and disruptive classroom behavior at school have been treated extensively in behavior modification and, hence, were selected as the initial focus of the case descriptions. Characteristics of the case such as intelligence, nature of the problem, setting in which the program was conducted, and person who sought treatment were intentionally varied to ensure that treatment evaluations were not restricted to unique characteristics of the descriptions.

Treatment Conditions

After hearing one of the cases, the students heard a description of each of the four treatments. The treatments were described in such a way that were specific to the previous case that the student heard. Hence, the treatment was described as it had actually been conducted in the home or at school. Four different treatments were described.

Reinforcement. The description of reinforcement emphasized positive reinforcement for behaviors that were incompatible with the undesired behaviors described in the case. For example, for Ann, parents provided stars on a special chart (tokens) for complying with instructions (e.g., getting dressed, going to bed on time) and behaving cooperatively. Special privileges (helping her mother bake, watching TV, buying toys) were used as back-up events. A similar program was described for Ralph, with the necessary changes being made for the specific contingencies and back-up events.

Time out from reinforcement. Time out consisted of punishing directly the undesired behaviors described for each child. Time out consisted of 10 minutes of isolation for each occurrence of an undesired behavior. For example, for Ralph, the teacher placed the child in an isolated

part of the classroom that was partitioned off for this purpose. After the allotted time out period elapsed, Ralph was allowed to return to his work. For Ann, time out consisted of sitting in a corner of a room in her house away from others for 10 minutes.

Drug treatment. Drug treatment was described as the administration of a stimulant [methylphenidate (Ritalin)] which was designed to suppress inappropriate behavior. The treatment was described as one which slows motor activity and is of use in controlling disruptive behavior. Administration by the parents or by the teacher (through parental request) for Ann and Ralph, respectively, was said to be three times per day in capsule form and based upon body weight to determine optimal dose.

Electric shock. Shock was described as being delivered for instances of inappropriate behavior. A portable battery-operated shock prod was described which could be easily carried by the parent or teacher and was touched to the child's arm or leg for inappropriate behavior. The shock was said to be very quick and only moderately painful.

As noted earlier, the four treatments were described in a different order to meet the requirements of a 4×4 Latin Square. Each of the treatments and their descriptions was derived from versions reported in the literature. Hence, drug doses, shock apparatus, reinforcement contingencies, and isolation all entailed procedures that have been used in actual treatment.

It should be noted that although each of the treatments constitutes actual procedures used in the applied literature, each was *not necessarily* used for the target behaviors described for deviant problematic behavior. For example, shock and drug treatment are not applied to mildly disruptive behavior. However, these procedures were included for two reasons. First, one would expect that as more severe forms of treatment, shock and drug treatment would be less acceptable than would reinforcement and time out. Hence, inclusion of these procedures might provide evidence that the Treatment Evaluation

Inventory can discriminate acceptability of different procedures. Second, in the second experiment, these interventions *were* applied to target problems of much greater client severity and hence are more viable as treatment alternatives.

RESULTS

Treatment Evaluations

To examine participant evaluations of treatment, for each of the dependent measures, separate 4×4 Replicated Latin Square analyses of variance were completed. For the Treatment Evaluation Inventory and each of the three Semantic Differential scales, neither the Sequence in which treatments were presented nor the Order in which a treatment appeared were statistically significant. However, the effects of Treatment Conditions were significant for total acceptability scores from the Treatment Evaluation Inventory, $F(3, 6) = 63.53$, $p < .001$; and for the Semantic Differential scales measuring Evaluative, $F(3, 6) = 44.72$, $p < .001$; Potency, $F(3, 252) = 22.77$, $p < .001$; and Activity dimensions, $F(3, 252) = 35.86$, $p < .001$.⁴

The analyses of variance indicated statistically significant effects for Treatment Conditions. To evaluate the strength of the relationship between treatments and acceptability ratings, w^2 (omega squared) was computed (Kirk, 1968). For the Treatment Evaluation Inventory and Evaluative scale of the Semantic Differential, 61 and 59%

of the variance in acceptability ratings was accounted for, respectively. Hence, the strength of the treatment effect was relatively large. For the Potency and Activity scales, the variance accounted for was 15 and 24%, respectively.

To evaluate the sources of difference among treatment conditions, Newman-Keuls comparisons were performed. The means for treatment conditions are presented in Figure 1. The results of the multiple comparisons, presented in Table 1, indicate that reinforcement was rated as significantly more acceptable than any of the other procedures on the Treatment Evaluation Inventory. In turn, time out, drug therapy, and shock were less acceptable alternatives. The differences among all of the means were statistically significant indicating that treatment alternatives were clearly distinguished.

Multiple comparisons of conditions for the Semantic Differential scales also distinguished groups (see Table 1). For the Evaluative scale, reinforcement was rated more positively than the other treatments, followed in order by time out, drug therapy, and shock (see Figure 1). Again, each of the means was significantly different from the others. For the Potency scale, the results differed slightly. Reinforcement was rated as the least powerful followed by time out, drug therapy, and shock. For the Activity scale, the shock procedure was rated as significantly

Table 1
Newman-Keuls Comparisons of Treatment Conditions

<i>Dependent Measures</i>	<i>Conditions</i>			
<i>Treatment Evaluation Inventory</i>				
Acceptability Total	R	T	D	S
<i>Semantic Differential Ratings</i>				
Evaluative	R	T	D	S
Potency	S	D	T	R
Activity	S	D	R	T

Note. Any two means underlined by the same line are not significantly different, whereas any two means not underlined by the same line are significantly different. All differences are at the $p < .05$ level. R = reinforcement for incompatible behavior, D = drug therapy, S = shock, T = time out from reinforcement.

⁴The Replicated Latin Square analysis yields two error terms, occasionally referred to as Latin Square error (or Square Uniqueness) and the residual. The residual term usually is considered as the appropriate error term to test Treatment Effects unless it is smaller than Latin Square error. When this latter term is larger, it is recommended as the error term in place of the residual (Myers, 1966). The Latin Square error term provides a negatively biased and, hence, highly conservative F test both because of the magnitude of the mean square and the lower degrees of freedom in the denominator for the F test. The different degrees of freedom for the tests of Treatments in the analyses reflect selection of the more conservative test.

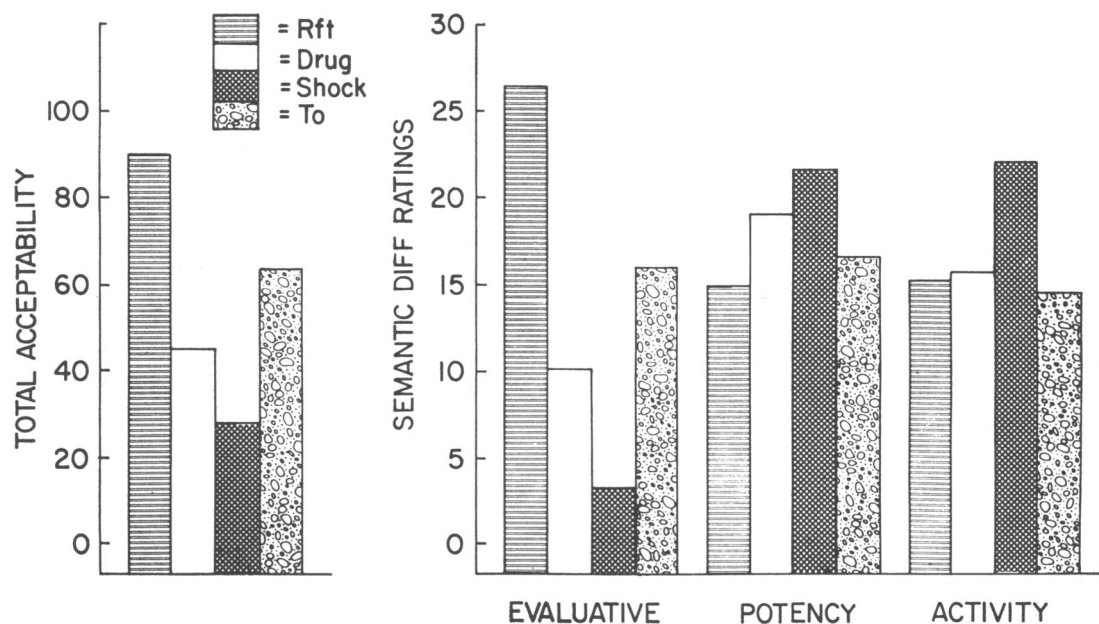


Fig. 1. Mean acceptability ratings for the Treatment Evaluation Inventory (total acceptability) and Semantic Differential Scales (evaluative, potency, and activity) for each treatment condition.

more active than all other treatments, which were not different from each other.

Supplementary Analyses

As noted earlier, students rated different treatments as applied to one of two clinical case descriptions (Ann or Ralph). The cases differed in gender, type of behavioral problem, and locus of treatment (home or school), to evaluate whether the results might be restricted to a single type of treatment application. Analyses of variance were completed to examine whether students rated the treatments differently as a function of the case description. A 2×4 (Child \times Treatment Condition) analysis of variance was performed for each dependent variable. In each analysis, Treatment Conditions was a significant effect which, of course, is redundant with the previous Latin Square analyses. Child case was not a significant factor nor was the interaction of Child \times Treatment conditions in any of the analyses. Thus, the ratings that students provided for the different treatments did not depend upon which case description they had received.

The possibility exists within the study that male and female participants might evaluate treatment differently. A sufficient number of participants was not available to include participant Gender as a separate factor in the overall analysis. Hence, separate analyses of variance were completed to evaluate whether each treatment condition was evaluated differently by females and males. Among the 16 dependent measures (four measures for each of four treatments), no significant differences were obtained. Thus, the ratings of treatment did not depend upon student gender.

EXPERIMENT 2

The previous experiment demonstrated that participants readily distinguished treatments in overall acceptability and the evaluative dimension of the Semantic Differential. Although differences were marked, the evaluations may have resulted in part because of the specific clinical problems to which they were applied. Each of the case descriptions presented a child with disruptive, uncooperative, or potentially dangerous

behavior. The problems were selected because of the relatively high proportion of behavioral interventions with children who have such behaviors (Kazdin, 1975). Yet, it is likely the cases might be considered to present relatively moderate clinical problems. Although some of the behaviors in need of treatment were potentially dangerous, each case was not actually dangerous nor represented severe psychopathology or behavioral disorders. Treatments may be differentially acceptable depending upon the problems to which they are applied. With more severe problems, such intrusive treatments as shock and drug therapy may not be evaluated as negatively. To evaluate the role of the severity of the client's problem, the present study included cases varying in severity of the presenting problem.

Overview

Undergraduate students rated the different treatment conditions included in Experiment 1. The four different treatments were presented in a 4×4 Replicated Latin Square. Instead of two different clinical cases, four cases were used as the stimulus material. Students rated the different treatments as they applied to one case. The cases consisted of two girls and two boys. The case descriptions within gender categories differed in whether the problem they presented in treatment was relatively mild or severe, as described below.

METHOD

Participants and Procedures

Participants consisted of 68 females and 26 males who were drawn from undergraduate courses in psychology, education, and human development. All participants received course credit for their participation. Students were assigned to one of 16 conditions (based upon the case, severity, and sequence of treatments).

The procedures were identical to those described in Experiment 1 in which each student heard the case description and rated the four different treatments that were presented consecu-

tively in prearranged sequences, as specified by the Latin Square. The assessment devices were identical to those used in Experiment 1.

Treatment Conditions

The treatments that students rated consisted of reinforcement of incompatible behavior, time out from reinforcement, drug therapy, and electric shock. Four cases were included as stimulus material, two descriptions of a 5-year-old girl named Ann and two descriptions of a 10-year-old boy named Ralph. The descriptions for a given child were consistent in terms of background information (e.g., living with parents and a sibling). The cases differed in the severity of the problems that they presented. For both Ann and Ralph, the relatively *moderate* descriptions included the problems described in Experiment 1. Thus, Ann's primary problem was noncompliance, parental defiance, whining, and problems in sibling interaction. Analogously, Ralph's problems included disruptive classroom behavior, being out of his seat, consistently poor concentration, and not doing his classwork.

For the *severe* descriptions of Ann and Ralph, the clinical problems changed. Ann was said to be very aggressive at home as well as noncomplaint. Her aggressiveness included physically dangerous acts such as self-injurious behavior (headbanging with her fist and small objects), physically fighting with her younger brother, and throwing objects at her brother and parents. She was credited with tantrums that involved throwing objects around the house which often resulted in broken items. Her parents were described as very concerned about the permanent damage she might do to herself and to her brother and the social isolation that her aggressive behavior had caused with her peers.

In the severe description, Ralph was described as being in a special education classroom in a state institution (rather than in an ordinary elementary school in the moderate case description). Ralph was credited with hyperactivity as well as more extremely disruptive behavior than in the moderate condition by running around the

room and performing bizarre behaviors such as hitting himself hard. To add to the severity, Ralph was said to have a history of severely disruptive behavior and that he had been recalcitrant to previous treatments. The teacher was noted as seeking treatment because Ralph could not be controlled in the classroom.

In general, background material (e.g., age, number of siblings, home environment) within each the two descriptions of a given case was constant. However, the descriptions differed in the severity of the clinical problem that was presented, its chronicity, and dangerous to the client and to others. For the case descriptions of Ralph, additional cues were provided to convey differences in severity of the problem including differences in level of retardation (IQ of 40 vs. 70), and place of treatment (institution vs. ordinary elementary school) for severe and moderate conditions, respectively.

RESULTS

Treatment Evaluations

To evaluate Treatment Conditions, separate 4×4 Replicated Latin Square analyses of variance were completed for the Treatment Evaluation Inventory and Semantic Differential scales. Neither Sequence nor Order effects were significant in the analyses. However, Treatment Condition was significant for total acceptability scores on the Treatment Evaluation Inventory, $F(3, 6) = 64.64, p < .001$; and for the Semantic Differential scales measuring Evaluative, $F(3, 6) = 76.64, p < .001$; Potency, $F(3, 6) = 11.02, p < .01$; and Activity dimensions, $F(3, 6) = 24.79, p < .001$.

As in the previous Experiment, the strength of the relationship between Treatment Conditions and acceptability ratings was relatively large. Calculation of w^2 (omega squared) indicated that a substantial portion of variance for the Treatment Evaluation Inventory and Evaluative scale was accounted for by the different treatments (67 and 64% of the variance, respectively). For the Potency and Activity scales,

the strength was lower (18 and 33% of the variance, respectively).

To evaluate the sources of difference among treatment means (see Fig. 2), Newman-Keuls comparisons were performed. The results of the multiple comparisons (see Table 2) showed that reinforcement was rated as significantly more acceptable on the Treatment Evaluation Inventory than the other treatment groups. In contrast, shock was significantly less acceptable than the other treatments. Drug therapy and time out, which did not differ in acceptability from each other, were different from shock and reinforcement. This pattern of means was identical to the pattern obtained for the Evaluative Scale of the Semantic Differential (see Table 2). For the Potency scale, shock was rated as more powerful than all other treatments which did not differ from each other. This pattern was also obtained for the Activity scale (see Table 2).

Severity of Clinical Case

A major purpose of Experiment 2 was to evaluate whether students evaluated treatments differently as a function of the severity of the case to which treatment was applied. Presumably, more intrusive treatments might be more acceptable if the clinical problem is relatively severe. To evaluate severity of the clinical problem and whether the case descriptions evoked

Table 2
Newman-Keuls Comparisons of Treatment Conditions

Dependent Measures	Conditions			
<i>Treatment Evaluation Inventory</i>				
Acceptability Total	R	<u>T</u>	<u>D</u>	S
<i>Semantic Differential Ratings</i>				
Evaluative	R	<u>T</u>	<u>D</u>	S
Potency	S	<u>T</u>	<u>D</u>	R
Activity	S	<u>R</u>	<u>T</u>	D

Note. Any two means underlined by the same line are not significantly different, whereas any two means not underlined by the same line are significantly different. All differences are at the $p < .05$ level. R = reinforcement for incompatible behavior, D = drug therapy, S = shock, T = time out from reinforcement.

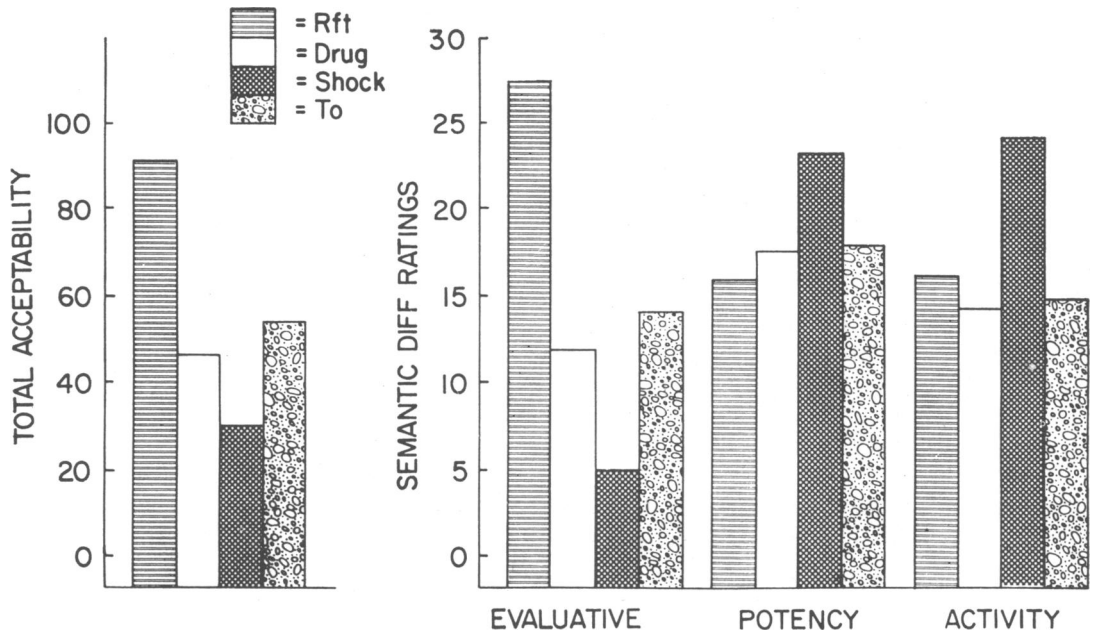


Fig. 2. Mean acceptability ratings for the Treatment Evaluation Inventory (total acceptability) and the Semantic Differential Scales (evaluative, potency, and activity) for each treatment condition.

differential reactions, Case Severity \times Child \times Treatment Condition ($2 \times 2 \times 4$) analyses of variance were completed.

For total acceptability scores on the Treatment Evaluation Inventory, a significant effect of Case Severity, $F(1, 108) = 7.03$, $p < .01$, was obtained. As expected, students rated treatments as significantly more acceptable for more severe case descriptions than for less severe case descriptions. Interestingly, the interaction of Case Severity \times Treatment was not significant. Hence, even though severity of the presenting problem evoked different acceptability scores, ordering of treatments and overall evaluations were not differentially affected.⁵

Although not of direct interest, it is worth mentioning that treatments were rated as more acceptable when applied to Ralph than when applied to Ann, $F(1, 108) = 4.71$, $p < .05$. These cases were intentionally varied along several dimensions including age, gender, and setting where treatment was conducted to maxi-

mize the stimulus sampling of case differences (Maher, 1978). Hence which of the above, or any other factors, that accounted for differences in acceptability between the case descriptions may be of interest for future research. For present purposes, the important point is that neither child case nor severity of the case interacted with treatment acceptability.

The results were similar for the Evaluative scale of the Semantic Differential showing a significant effect of Child, $F(1, 108) = 6.75$, $p < .01$; and an effect that approached significance for Case Severity, $F(1, 108) = 3.73$, $p < .06$. Treatments were rated as more acceptable for severe case descriptions and for descriptions of Ralph. Although Treatment Condition was a significant effect in the analysis, it did not interact with the other variables. For Potency and Activity, neither Child nor Case Severity yielded significant effects.

Although Child and Case Severity achieved statistically significant effects on total acceptability scores and Evaluative ratings on the Semantic Differential, the magnitude of these effects was weak. Calculation for w^2 (omega

⁵The effect of treatment conditions was also significant but will not be evaluated here because of its redundancy with the primary Latin Square analyses.

squared) for each of the statistically significant effects revealed that less than 1% of the acceptability ratings was accounted for by the Child or Case Severity. Hence, the particular child who was described and the severity of the presenting clinical problem exerted relatively little impact on overall acceptability ratings.

Supplementary Analyses

Given the differential acceptability ratings based upon Child and Severity of the clinical case, it might be of interest to evaluate whether gender of the participants contributed to treatment evaluation. An insufficient number of male students precluded including Participant Gender in the previous analyses as a completely crossed factor. However, separate analyses were completed for Participant Gender to evaluate whether treatments were evaluated differently. Analyses for each of the four dependent variables for the four different treatment conditions yielded a significant effect for acceptability ratings of reinforcement, $F(1, 110) = 8.54, p < .01$; and acceptability ratings of shock, $F(1, 110) = 3.92, p < .05$. Males rated reinforcement as less acceptable than did females and rated shock as more acceptable than did females. However, the ordering of treatments among male and female participants and the pattern of significance among different treatments for males and for females did not differ. Hence, the different ratings made by male and female students do not alter the conclusions reached earlier. Further, the actual amount of variance accounted for by Participant Gender was small. For reinforcement and shock treatments, only 6 and 3% of the variance in acceptability ratings, respectively, was accounted for by the participant's sex.

DISCUSSION

The major findings of the present investigations were: (1) Students readily distinguished the acceptability of alternative treatment techniques as applied to child behavior problems. (2) Among the alternative procedures, rein-

forcement of incompatible behavior was evaluated as the most acceptable treatment, followed, in order, by time out from reinforcement, drug therapy, and electric shock. (3) Severity of the clinical problems to which the various treatments were applied influenced the overall rated acceptability of treatment, with all treatments rated as more acceptable for more severe clinical cases.

The main findings were the markedly different acceptability evaluations for the different treatments. Treatments were consistently different from each other on both acceptability ratings for the Treatment Evaluation Inventory and the Evaluative dimension of the Semantic Differential. Although severity of the clinical case to which treatments were applied was related to treatment acceptability, this variable did not alter the pattern of acceptability among the treatments, exerted relatively little influence on the magnitude of rated acceptability, and accounted for little variance in acceptability evaluations. In contrast, a large portion of variance of the acceptability ratings was consistently accounted for by the different treatment procedures.

Although treatments were distinct on acceptability and evaluative ratings, they were not distinguished as readily on the Potency and Activity dimensions of the Semantic Differential. On these latter measures, shock was consistently rated as the most potent and active treatment but other treatments generally did not differ from each other. Two aspects of the overall findings are of particular interest. Overall, shock was seen as a relatively unacceptable treatment and as a very strong intervention. Although reinforcement, time out, and drug therapy varied considerably in acceptability, they were generally viewed as similar in overall potency and activity.

The present investigations represent an initial attempt to evaluate acceptability of treatment procedures. Although treatments were clearly distinguished, several limitations are important to consider when interpreting the findings. To

begin with, only four different treatments were evaluated. Several other procedures that might be used for the target problems described in the cases such as response cost, positive practice, and physical restraint were not included. Moreover, among the treatments that were included, many different variations were not investigated. For example, time out from reinforcement in the present study consisted of 10 minutes of isolation from the situation at home or in the classroom. However, time out can vary considerably in how it is conducted, the duration of time out, whether the individual is removed from the situation at all, whether others are removed from the presence of the target child, and whether a period of time of not earning reinforcers is simply set aside in the situation (e.g., Hobbs & Forehand, 1977). Presumably, various forms of time out differ greatly in their acceptability (cf. Foxx & Shapiro, 1978). Similarly, with other procedures, technique variations may be evaluated somewhat differently from the specific techniques used in the present studies.

Aside from a restricted set of treatments, other conditions of the present investigation warrant comment. Evaluation of acceptability was examined in the context of behavior problem children. It would be of interest to evaluate a larger number of cases, varying in age and type of clinical problem. Also, evaluation of treatment may depend upon a variety of factors such as whether alternative and less intrusive treatments had been unsuccessfully applied in the history of the case, the degree of danger the clients presented to themselves or to others, the ease of treatment administration, and the potential risks and benefits of treatment. Two major determinants of acceptability of treatment may be the relative effectiveness of the procedure and the ease of its implementation in applied settings. For example, parents and teachers may view a particular procedure as a more viable alternative than another procedure on the basis of practical considerations. The above variables were beyond the scope of the present investigations but might well influence how potential

consumers evaluate treatment. Additional investigations about how consumers evaluate treatment and whether such evaluations influence utilization of the procedures seem warranted.

Finally, the participants in the present investigations were undergraduate college students. Although many such individuals are potential consumers of treatment, they had not solicited treatment for their children. It is quite possible that acceptability of treatment may vary among individuals whose problem children have led them to seek consultation. On the other hand, the present group represents well educated individuals whose views of alternative treatments may be less negative than less informed individuals. It is important to evaluate the acceptability of treatments among those individuals who actually seek treatment.

The main criteria for evaluating intervention procedures are the various measures of outcome including the magnitude, durability, and breadth of treatment effects, and evaluations of treatment effects by the clients and others with whom they may interact. Among effective treatment alternatives, several other dimensions become relevant including cost of the procedures, ease of disseminating treatments to the public, and ease of administration (Kazdin & Wilson, 1978). Acceptability of treatment is one of the important dimensions in its own right that may determine whether a treatment, once shown to be effective, will be utilized by the public.

Researchers have become increasingly sensitive to facets that may contribute to the utilization of treatment by developing techniques that can be implemented with little or no contact with a therapist or consultant (Glasgow & Rosen, 1978). For example, for many clients, conducting interventions on their own may make the technique more palatable and, hence, more readily utilized. Of course, many other treatments are not utilized as self-help procedures and yet may be differentially attractive for a variety of reasons. As evidence accumulates about the efficacy of alternative procedures, additional work might be worthwhile that assesses

those factors that influence client selection of and adherence to treatment. As suggested by the present investigations, characteristics of the procedures themselves may account for a major component of client evaluations of treatment.

The present investigations represent an initial step to evaluate alternative treatments and the variables that contribute to their acceptability. A long-term goal of such research is not merely to catalogue different client reactions. Rather, the purpose is to evaluate those variables that may influence evaluations of potential consumers so that highly effective procedures may, if necessary, be altered in light of these variables. For many effective procedures, ancillary or nonessential components may be alterable to make them more acceptable with little or no loss of efficacy.

REFERENCES

- Budd, K. S., & Baer, D. M. Behavior modification and the law: Implications of recent judicial decisions. *Journal of Psychiatry and Law*, 1976, 4, 171-244.
- Campbell, D. T., & Fiske, D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Crowe, M. J., Marks, I. M., Agras, W. S., & Leitenberg, H. Time-limited desensitisation, implosion and shaping for phobic patients: A crossover study. *Behaviour Research and Therapy*, 1972, 10, 319-328.
- Drabman, R., Spitalnik, R., & Spitalnik, K. Sociometric and disruptive behavior as a function of four types of token reinforcement programs. *Journal of Applied Behavior Analysis*, 1974, 7, 93-101.
- Foxx, R. M., & Azrin, N. H. Restitution: A method of eliminating aggressive-disruptive behavior of retarded and brain damaged patients. *Behaviour Research and Therapy*, 1972, 10, 15-27.
- Foxx, R. M., & Azrin, N. H. Dry pants: A rapid method of toilet training children. *Behaviour Research and Therapy*, 1973, 11, 435-442.
- Foxx, R. M., & Shapiro, S. T. The timeout ribbon: A nonexclusionary timeout procedure. *Journal of Applied Behavior Analysis*, 1978, 11, 125-136.
- Garfield, S. L. Research problems in clinical diagnosis. *Journal of Consulting and Clinical Psychology*, 1978, 46, 596-607.
- Glasgow, R. E., & Rosen, G. M. Behavioral bibliography: A review of self-help behavior therapy manuals. *Psychological Bulletin*, 1978, 85, 1-23.
- Harman, H. H. *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press, 1976.
- Hobbs, S. A., & Forehand, R. Important parameters in the use of timeout with children: A re-examination. *Journal of Behavior Therapy and Experimental Psychiatry*, 1977, 8, 365-370.
- Kazdin, A. E. Characteristics and trends in applied behavior analysis. *Journal of Applied Behavior Analysis*, 1975, 8, 332.
- Kazdin, A. E. *History of behavior modification: Experimental foundations of contemporary research*. Baltimore: University Park Press, 1978.
- Kazdin, A. E., & Wilson, G. T. *Evaluation of behavior therapy: Issues evidence, and research strategies*. Cambridge, Mass.: Ballinger, 1978.
- Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Cal.: Brooks/Cole, 1968.
- Leitenberg, H. Behavioral approaches to treatment of neuroses. In H. Leitenberg (Ed.), *Handbook of behavior modification and behavior therapy*. Englewood Cliffs, N.J.: Prentice-Hall, 1976.
- Liberman, R. P., Levine, J., Wheeler, E., Sanders, N., & Wallace, C. J. Marital therapy in groups: A comparative evaluation of behavioral and interactional formats. *Acta Psychiatrica Scandinavica*, 1976, Supplementum, 266.
- Maher, B. A. Stimulus sampling in clinical research: Representative design reviewed. *Journal of Consulting and Clinical Psychology*, 1978, 46, 643-647.
- Marks, I. Behavioral psychotherapy of adult neuroses. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis* (2nd edition). New York: Wiley, 1978.
- Martin, R. *Legal challenges to behavior modification: Trends in schools, corrections and mental health*. Champaign, Ill.: Research Press, 1975.
- May, J. G., Risley, T. R., Twardosz, S., Friedman, P., Bijou, S. W., & Wexler, D. *Guidelines for the use of behavioral procedures in state programs for retarded persons*. Arlington, Tex.: National Association for Retarded Citizens, 1976.
- Myers, J. L. *Fundamentals of experimental design*. Boston: Allyn and Bacon, 1966.
- O'Leary, S. G., & O'Leary, K. D. Behavior modification in the school. In H. Leitenberg (Ed.), *Handbook of behavior modification and behavior therapy*. Englewood Cliffs, N.J.: Prentice-Hall, 1976.
- Osgood, C. E., Suci, G. J., & Tannebaum, P. H. *Measurement of meaning*. Urbana: University of Illinois Press, 1957.
- Porterfield, J. K., Herbert-Jackson, E., & Risley, T. R. Contingent observation: An effective and acceptable procedure for reducing disruptive behavior

- of young children in a group setting. *Journal of Applied Behavior Analysis*, 1976, 9, 55-64.
- Rosenbaum, A., O'Leary, K. D., & Jacob, R. G. Behavioral intervention with hyperactive children: Group consequences as a supplement to individual contingencies. *Behavior Therapy*, 1975, 6, 315-323.
- Strupp, H. H., & Hadley, S. W. A tripartite model of mental health and therapeutic outcomes. *American Psychologist*, 1977, 32, 187-196.
- Sulzer-Azaroff, B., & Mayer, G. R. *Applying behavior-analysis procedures with children and youth*. New York: Holt, Rinehart & Winston, 1977.
- Webster, D. R., & Azrin, N. H. Required relaxation: A method of inhibiting agitative-disruptive behavior of retardates. *Behaviour Research and Therapy*, 1973, 11, 77-78.
- Wolf, M. M. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 1978, 11, 203-214.

Received March 13, 1979

Final acceptance September 20, 1979